In [1]:

```python
import requests
from bs4 import BeautifulSoup
import pandas as pd
```

In [2]:

```python
url = 'https://www.themoviedb.org/movie?page='
page_url = 'https://www.themoviedb.org'
```

In [3]:

```python
content=requests.get(url,headers={'User-Agent':'Mozilla/5.0'}).text
```

In [4]:

```python
## movie_url of 1st page
soup = BeautifulSoup(content,'lxml')
movie_url_1st_page = []
movie_name_lst = []
data = soup.find_all('div',class_ = 'card style_1')
for i in data:
    movie_code = i.a['href']
    movie_url_1st_page.append(page_url+movie_code)
    names = i.a['title']
    movie_name_lst.append(names)
```

In [5]:

```
director_lst = []
gen_lst = []
run_time = []
release_lst = []
raiting_lst = []

for link in movie_url_1st_page:
    content=requests.get(link,headers={'User-Agent':'Mozilla/5.0'}).text

    soup = BeautifulSoup(content,'lxml')
    data = soup.find_all('div',class_ = 'header_poster_wrapper true')

    raiting = soup.find('div',class_ = 'user_score_chart')['data-percent']
    raiting_lst.append(raiting)


    release_date = soup.find('span',class_ = 'release').text.split()[0]
    release_lst.append(release_date)

    director = soup.find('li',class_ = 'profile').a.text
    director_lst.append(director)

    val = str(soup.find('span',class_ = 'genres').text)
    genres = val.replace('\n','')
    gen_lst.append(genres)


    runtime = str(soup.find('span',class_ = 'runtime').text)
    time = runtime.replace('\n','')
    run_time.append(time)

    movie_data_dic = {
        'Movie Name': movie_name_lst,
        'Raiting' : raiting_lst,
        'Release Date' : release_lst,
        'Run Time': run_time,
        'Genres': gen_lst,
        'Director': director_lst,
        'Movie_link': movie_url_1st_page

     }
```

In [6]:

```
Df = pd.DataFrame(movie_data_dic)
```

In [7]:

```
Df
```

Out[7]:

	Movie Name	Raiting	Release Date	Run Time	Genres	Director	
0	Doctor Strange in the Multiverse of Madness	75.0	05/06/2022	2h 6m	Fantasy, Action, Adventure	Steve Ditko	http
1	Fantastic Beasts: The Secrets of Dumbledore	68.0	04/15/2022	2h 22m	Fantasy, Adventure, Action	David Yates	http
2	Dog	74.0	02/18/2022	1h 42m	Drama, Comedy	Reid Carolin	http
3	Sonic the Hedgehog 2	77.0	04/08/2022	2h 2m	Action, Adventure, Family, Comedy	Josh Miller	http
4	Morbius	65.0	04/01/2022	1h 45m	Action, Science Fiction, Fantasy	Daniel Espinosa	http
5	The Lost City	68.0	03/25/2022	1h 52m	Action, Adventure, Comedy	Adam Nee	http
6	Spider-Man: No Way Home	81.0	12/17/2021	2h 28m	Action, Adventure, Science Fiction	Steve Ditko	http
7	Memory	73.0	04/29/2022	1h 54m	Action, Thriller, Crime	Martin Campbell	http
8	Collision	60.0	06/16/2022	1h 39m	Thriller, Crime, Drama	Fabien Martorell	http
9	Centauro	66.0	06/15/2022	1h 29m	Action, Crime, Thriller	Daniel Calparsoro	http
10	Spiderhead	57.0	06/17/2022	1h 46m	Science Fiction, Thriller	Joseph Kosinski	http
11	Jurassic World Dominion	67.0	06/10/2022	2h 27m	Action, Adventure, Science Fiction	Colin Trevorrow	http
12	The Black Phone	72.0	06/24/2022	1h 43m	Horror, Thriller	Scott Derrickson	http
13	Shark Bait	70.0	05/13/2022	1h 27m	Horror, Thriller, Action	James Nunn	http
14	Panama	59.0	03/17/2022	1h 40m	Action, Thriller	Mark Neveldine	http
15	The Northman	73.0	04/22/2022	2h 17m	Action, Adventure, Fantasy	Robert Eggers	http
16	Turning Red	75.0	03/10/2022	1h 40m	Animation, Family, Comedy, Fantasy	Domee Shi	http
17	Uncharted	71.0	02/18/2022	1h 56m	Action, Adventure	Ruben Fleischer	http
18	The Desperate Hour	61.0	09/12/2021	1h 24m	Thriller	Phillip Noyce	http

	Movie Name	Raiting	Release Date	Run Time	Genres	Director	
19	Hustle	79.0	06/03/2022	1h 58m	Drama, Comedy	Jeremiah Zagar	http

In [ ]:

```

```

In [8]:

```python
url_lst = []
for u in range(0,501):
    url_lst.append(url+str(u))
```

In [ ]:

```

```

In [9]:

```python
movie_url_all_pages = []
movie_name_lst = []

for link in url_lst:
    content=requests.get(link,headers={'User-Agent':'Mozilla/5.0'}).text
    soup = BeautifulSoup(content,'lxml')
    data = soup.find_all('div',class_ = 'card style_1')

    for i in data:
        movie_code = i.a['href']
        movie_url_all_pages.append(page_url+movie_code)
        names = i.a['title']
        movie_name_lst.append(names)

```

In [10]:

```python
len(movie_name_lst)
```

Out[10]:

```
10000
```

In [33]:

```
director_lst = []
gen_lst = []
run_time = []
release_lst = []
raiting_lst = []

for link in movie_url_all_pages:
    content=requests.get(link,headers={'User-Agent':'Mozilla/5.0'}).text


    soup = BeautifulSoup(content,'lxml')
    data = soup.find_all('div',class_ = 'header_poster_wrapper true')

    raiting = soup.find('div',class_ = 'user_score_chart')['data-percent']
    raiting_lst.append(raiting)


    release_date = soup.find('span',class_ = 'release').text.split()[0]
    release_lst.append(release_date)

    director = soup.find('li',class_ = 'profile')
    if director is not None:
            director=(director.p.text)
    director_lst.append(director)

    val = str(soup.find('span',class_ = 'genres').text)
    genres = val.replace('\n','')
    gen_lst.append(genres)

    runtime = soup.find('span',class_='runtime')
    if runtime is not None:
        runtime=(runtime.text.strip())
        #run_time.append(runtime)

    run_time.append(runtime)

    movie_data_dic = {
        'Movie Name': movie_name_lst,
        'Raiting' : raiting_lst,
        'Release Date' : release_lst,
        'Run Time': run_time,
        'Genres': gen_lst,
        'Director': director_lst,
        'Movie_link': movie_url_all_pages

     }
```

In [34]:

```
len(run_time)
```

Out[34]:

```
10000
```

In [35]:

```
len(director_lst)
```

Out[35]:

```
10000
```

In [36]:

```
len(movie_url_all_pages)
```

Out[36]:

```
10000
```

In [37]:

```
len(release_lst)
```

Out[37]:

```
10000
```

In [38]:

```
df = pd.DataFrame(movie_data_dic)
df
```

Out[38]:

	Movie Name	Raiting	Release Date	Run Time	Genres	Director	
0	Doctor Strange in the Multiverse of Madness	75.0	05/06/2022	2h 6m	Fantasy, Action, Adventure	Steve Ditko	https://v
1	Fantastic Beasts: The Secrets of Dumbledore	68.0	04/15/2022	2h 22m	Fantasy, Adventure, Action	David Yates	https://v
2	Dog	74.0	02/18/2022	1h 42m	Drama, Comedy	Reid Carolin	https://v
3	Sonic the Hedgehog 2	77.0	04/08/2022	2h 2m	Action, Adventure, Family, Comedy	Josh Miller	https://v
4	Morbius	65.0	04/01/2022	1h 45m	Action, Science Fiction, Fantasy	Daniel Espinosa	https://v
...	...	...	...	...	...	...	
9995	The Scary House	57.0	10/30/2020	1h 40m	Family, Fantasy, Horror	Daniel Prochaska	https://v
9996	Maria Full of Grace	73.0	04/01/2004	1h 41m	Drama, Thriller, Crime	Joshua Marston	http
9997	The Art of Getting By	66.0	06/17/2011	1h 23m	Drama, Romance	Gavin Wiesen	https:/
9998	I'm a Girl, I'm a Princess	60.0	10/28/2021	2h	Drama	Federico Palazzo	https://v
9999	Lake Eerie	38.0	01/15/2016	1h 44m	Thriller, Horror, Science Fiction	Chris Majors	https://v

10000 rows × 7 columns

In [41]:

```
df.to_csv("Movie_Data.csv")
```

In [ ]:

```


```

In [ ]:

In [ ]:

In [ ]: